



Europäisches Patentamt
European Patent Office
Office européen des brevets



(11) Publication number: 0 607 054 A2

(19)

(12)

(21) Application number: 94300267.5

(22) Date of filing: 14.01.94

EUROPEAN PATENT APPLICATION

(51) Int. Cl.⁵: C12N 15/10, C12N 15/12,
C12N 15/62, C12N 15/85,
C07K 13/00, C07K 15/28,
C12Q 1/68, A61K 37/02,
A61K 39/395

(30) Priority: 14.01.93 JP 22098/93

(43) Date of publication of application:
20.07.94 Bulletin 94/29

(84) Designated Contracting States:
AT BE CH DE DK ES FR GB GR IE IT LI LU MC
NL PT SE

(71) Applicant: Honjo, Tasuku
Kanyu-chi,
Oiwake-cho
Kita-Shirakawa, Sakyo-ku, Kyoto (JP)

(71) Applicant: ONO PHARMACEUTICAL CO., LTD.
1-5, Doshomachi 2-chome
Chuo-ku Osaka 541 (JP)

(72) Inventor: Honjo, Tasuku Kanyu-chi

Oiwake-cho
Kita-Shirakawa
Sakyo-ku Kyoto (JP)
Inventor: Tashiro, Kei
93 Koyama Higashi
Ohno-cho

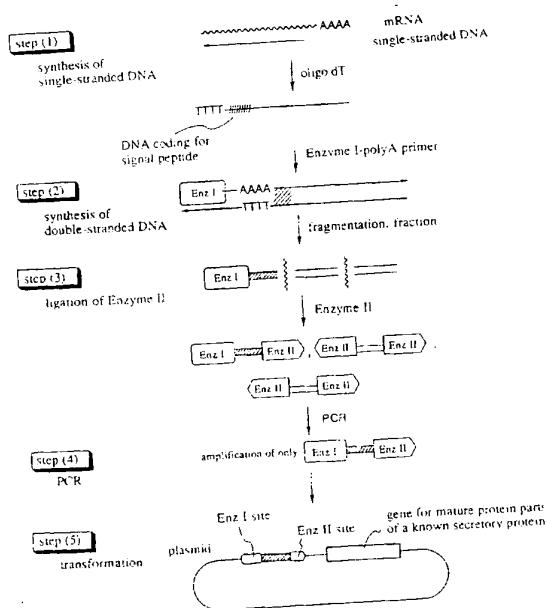
Kita-ku
Kyoto-shi Kyoto (JP)
Inventor: Tade, Hideaki c/o Minase Research
Institute
Ono Pharmaceuticals Co.Ltd.
3-1-1, Sakurai
Shimamoto-cho Mishima-gun Osaka (JP)

(74) Representative: Bentham, Stephen et al
J.A. KEMP & CO.
14 South Square
Gray's Inn
London WC1R 5LX (GB)

(54) Novel process for constructing a cDNA library and a novel polypeptide and DNA coding the same.

(57) A process for constructing a cDNA library which has a selectivity for signal peptides, that makes it possible to efficiently identify unknown and useful polypeptide comprising a signal peptide. A novel polypeptide consisting of 89 amino acids (including a signal peptide) produced by a stroma cell line, which is useful as an agent for preventing or treating, for example, anemia, leukopenia or infections and the like, and DNAs coding for said polypeptide, have been identified using the process of the invention.

Figure 1



EP 0 607 054 A2

Jouve, 18, rue Saint-Denis, 75001 PARIS

This invention relates to a process for constructing a cDNA library, and a novel polypeptide and a DNA coding for the same. More particularly, it relates to an efficient process for constructing a cDNA library having a high selectivity for signal peptides, and a novel polypeptide produced by a specific stroma cell line and a DNA coding for said polypeptide.

5 In order to obtain specific polypeptides (for example, proliferation and/or differentiation factors) or a DNA coding for the same, there have been generally employed methods comprising confirming the target biological activity in a tissue or a cell culture medium and then cloning of a gene through the isolation and purification of a polypeptide and further methods comprising expression-cloning of a gene with the guidance of the biological activity.

10 However, it is frequently observed that a gene, which has been cloned with the guidance of a certain activity, codes for a known polypeptide since many physiologically active polypeptides occurring *in vivo* have various biological activities. Further, most intravital physiologically active polypeptides are secreted only in a trace amount, which makes the isolation and purification thereof and the confirmation of biological activity difficult.

15 Recent rapid developments in techniques for constructing cDNAs and sequencing techniques have made it possible to quickly sequence a large amount of cDNAs. By utilising these techniques, a process, which comprises constructing cDNA libraries from various cells and tissues, cloning cDNAs at random, identifying the nucleotide sequences thereof, expressing the corresponding polypeptide and then analyzing its physiological functions, is now in use. Although this process is advantageous in that a gene can be cloned and information 20 on its nucleotide sequence can be obtained without any biochemical or genetic analysis, the target gene can often only be identified by chance.

25 The present inventors have studied the cloning of genes for proliferation and differentiation factors in hematopoietic and immune systems. They have paid attention to the fact that most secretory proteins such as proliferation and/or differentiation factors (for example various cytokines) and membrane proteins such as receptors thereof (hereinafter these proteins will be referred to generally as secretory proteins and the like) have sequences called signal peptides in the N-termini. Extensive studies have been conducted to provide a process for efficiently and selectively cloning genes coding for signal peptides. As a result, a process has now been devised whereby an N-terminal fragment can be efficiently amplified and the existence of a signal peptide can be easily examined.

30 In accordance with the present invention, cDNAs with a high probability of containing a signal peptide are ligated at both their ends to linkers containing restriction enzyme sites which are different from each other. These fragments alone are rapidly produced in a large amount by the polymerase chain reaction (PCR) method so as to elevate the content of the fragments with a high probability of containing a signal peptide. Next, the above-mentioned fragment is introduced into an expression vector containing DNA coding for a known secretory protein or the like but lacking DNA encoding the corresponding signal peptide. Many secretory proteins and the like are secreted or expressed on cell membrane even if the signal peptide has been substituted by a signal peptide of another protein. Therefore, if the known secretory protein or the like is expressed on cell membrane or outside the cells, this confirms that a cDNA fragment corresponding to a signal peptide has been introduced into the expression vector. Thus the inventors have devised a convenient method for detection of the insertion of cDNA encoding a signal peptide.

35 The polymerase chain reaction used in the present invention is known as a method for amplifying specific DNA fragments in a large amount. It is also known that many secretory proteins and the like can be expressed even if the signal peptide thereof is substituted by that of another secretory protein or the like. However, to the applicants' knowledge, there has been no suggestion that these techniques be used as the inventors have done to provide a process for selectively cloning a signal peptide.

40 It is known that hematopoietic cells secrete various proliferation and/or differentiation factors exemplified by interleukin.

45 The present invention further relates to a novel polypeptide obtained from hematopoietic cells and DNA coding for the same.

The inventors have sought a novel factor (polypeptide) produced by hematopoietic cells using the process which is the first subject of the present invention. As a result, a novel polypeptide and DNA coding for the same have been identified.

50 Computer searches to identify polypeptides having sequences identical or highly homologous with that of the polypeptide of the present invention and the DNAs coding for the same have not identified any such sequences. Thus, to the best of the applicants' knowledge, the polypeptide of the present invention and the DNA coding for the same are novel. Further, homology analysis has revealed that the polypeptide of the present invention is a member of the chemokine family as it has a pattern of Cys-X-Cys (X is optional amino acid).

55 Accordingly, the present invention provides a process for constructing a cDNA library comprising:

(1) synthesizing single-stranded DNA complementary to mRNA isolated from subject cells using a random primer and ligating oligo dT to the 3'-end of the single-stranded DNA;

(2) converting the single-stranded DNA obtained in (1) to a double-stranded DNA using as primer a poly

A oligomer ligated to a first restriction enzyme (enzyme I) site;

(3) fragmenting the double-stranded DNA obtained in (2), fractionating the fragments obtained by size,

ligating linker containing a second restriction enzyme (enzyme II) site, differing from the enzyme I, thereto

and fractionating;

(4) amplifying fragments containing the sites of restriction enzymes I and II by a polymerase chain reaction

using a first primer containing the enzyme I site and a second primer containing the enzyme II site, di-

gesting the cDNA thus amplified with the restriction enzymes, enzyme I and enzyme II and fractionating;

and

(5) ligating the cDNA fragment thus obtained upstream of a gene coding for a secretory protein or mem-

brane protein having deleted therefrom the sequence coding for a signal peptide, integrating the ligated

DNA into a eucaryotic cell expression plasmid vector, and transforming the vector into a host.

The invention further provides a polypeptide having the amino acid shown in SEQ ID No. 1, in substantially

purified form, a homologue thereof or a fragment of the sequence or homologue of a fragment, and DNA en-

coding such a polypeptide. The polypeptide having the sequence shown in SEQ ID No. 1 has been identified

using the process of the invention.

Fig. 1 is a conceptional view of the process for constructing a cDNA library according to the present in-

vention.

Fig. 2 is a flow chart for the construction of a plasmid vector pcDL-SR α .

Fig. 3 is a conceptional view of the process for constructing an EcoRI-SacI fragment of hG-CSF.

Fig. 4 is a conceptional view of the process for constructing an SacI-KpnI fragment of hTac cDNA.

Fig. 5 is a flow chart for the construction of pSGT and pSRT.

Fig. 6 is a conceptional view of the process for constructing an EcoRI-SacI fragment of hRAR α .

Fig. 7 is an FACS histogram showing the expression of a fusion protein hG-CSF-hTac on membrane.

Fig. 8 is an FACS histogram showing the expression of a fusion protein hRAR-hTac on membrane.

Fig. 9 is a conceptional view of the first half of the process for constructing the cDNA library of the Example.

Fig. 10 is a conceptional view of the second half of the process for constructing the cDNA library of the

Example.

Fig. 11 is a hydrophobicity profile of (a part of) the polypeptide according to the present invention.

The first subject of the present invention is concerned with a process for efficiently constructing a cDNA

library of signal peptides.

The process for constructing a cDNA library of signal peptides according to the present invention com-

prises the following steps:

(1) synthesizing a single-stranded DNA from mRNA isolated from the subject cells with the use of a random

primer and adding oligo dT to the 3'-end of the single-stranded DNA thus obtained;

(2) synthesizing a double-stranded DNA from the single-stranded DNA obtained in (1) using as a primer

a poly A oligomer ligated to a specific restriction enzyme (enzyme I) site;

(3) fragmenting the double-stranded DNA obtained in (2), fractionating the fragments obtained by size,

ligating linker containing a specific restriction enzyme (enzyme II) site, differing from the enzyme I, thereto

and fractionating again;

(4) performing a PCR using a first primer containing the enzyme I site and a second primer containing

the enzyme II site, digesting the cDNA thus amplified with enzyme I and enzyme II and fractionating; and

(5) ligating the cDNA fragment upstream of the gene for a known secretory protein or membrane protein

with the deletion of signal peptide and integrating the ligated DNA into an eucaryotic cell expression plas-

mid vector, followed by transformation.

Fig. 1 is a conceptional view of the above-mentioned steps.

Now each of these steps will be illustrated in more detail.

In the step (1), if required, the subject cells are stimulated with an appropriate stimulating agent, and then

the mRNA is isolated in accordance with known methods as described for example by Okayama H. et al., Meth-

ods in Enzymology, 154, 3 (1987).

As the subject cells, any cells may be used, so long as they have a possibility of producing a secretory

protein or the like. For example, nerve cells and hematopoietic cells may be cited therefor. A single-stranded

cDNA can be synthesized with the use of a random primer by methods known *per se*. A marketed random pri-

mer is available therefor. Subsequently, oligo dT is added to the 3'-end of the single-stranded cDNA by using

a terminal deoxytransferase.

In the step (2), a double-stranded cDNA is synthesized by methods known *per se*. Any restriction enzyme

sites may be used as the restriction enzyme (enzyme I) site to be ligated to a poly A oligomer serving as a primer and the restriction enzyme (enzyme II) site to be used in the next step (3), so long as they differ from each other. It is preferable to use EcoRI and SacI, as enzyme I and enzyme II respectively.

5 In the step (3), the double-stranded DNA is fragmented for example by ultrasonication so as to give an average cDNA length of 300 bp and the obtained fragments are fractionated into cDNAs of 200 to 500 bp by agarose gel electrophoresis (AGE). After blunting with T4DNA polymerase, enzyme II is ligated and the cDNAs are fractionated into DNAs of 200 to 500 bp again by agarose gel electrophoresis. As described above, any enzyme may be used as the enzyme II, so long as it differs from the enzyme I. The procedure in this step increases the likelihood that a cDNA fragment containing a signal peptide exists in the part located between the enzymes I and II.

10 15 In the step (4), PCR is carried out in order to further elevate the possibility that a cDNA fragment containing a signal peptide exists in the part located between the enzymes I and II. PCR is a well known technique and automated devices therefore are commercially available. It is sufficient to amplify 25 to 30 times. The cDNA thus amplified is digested with enzyme I and enzyme II and electrophoresed on an agarose gel to thereby fractionate into cDNAs of 200 to 500 bp.

20 15 In the step (5), a gene for a known secretory protein or the like with the deletion of signal peptide, which is called a reporter gene, and a cDNA fragment obtained in the above (4) are integrated into an eucaryotic cell expression plasmid vector in such a manner that the cDNA fragment is located upstream of the reporter gene. This is followed by transformation of the vector into a host.

25 Various known eucaryotic cell expression plasmid vectors for example, pcDL-SR α and pcEV-4 which are capable of acting in *Escherichia coli*, are usable in the present invention.

As the reporter gene, genes for mature protein parts of soluble secretory proteins and membrane proteins of any type are usable. The expression of these reporter genes may be confirmed by known methods such as antibody methods. Human IL-2 receptor α gene is especially suitable as a reporter gene.

25 30 A number of *E. coli* strains are known as hosts for transformation and any of these stains is usable. It is preferable to use DH5 competent cells [described in Gene, 96, 23 (1990)] therefor. Transformants may be incubated in a conventional manner and thus the cDNA library of the present invention can be obtained.

In the process for constructing a cDNA library according to the present invention, there is a high possibility that gene fragments coding for signal peptides are contained in the library. However, not every clone contains said fragment. Further, not all of the gene fragments code for unknown (novel) signal peptides. It is therefore necessary to screen a gene fragment coding for an unknown signal peptide from said library.

35 40 Namely, the cDNA library is divided into small pools of an appropriate size and integrated into an expression system. Examples of the expression system for producing a polypeptide include mammalian cells (for example, monkey COS-7 cells, Chinese hamster CHO cells, mouse L cells, etc.). Transfection may be performed in accordance with well known methods such as the DEAE-dextran method. After the completion of the incubation, the expression of the reporter gene is examined. It is known that a reporter gene would be expressed even though the signal peptide is the one characteristic to another secretory protein. That is to say, the fact that the reporter gene has been expressed indicates that a signal peptide of some secretory protein has been integrated into the library. Positive pools are further divided into smaller ones and the expression and the judgement are repeated until a single clone is obtained. The expression of the reporter gene can be judged by, for example, fluorescence-labeled antibody assay, enzyme-linked immunosorbent assay (ELISA) or radioimmunoassay (RIA), depending on kinds of the employed reporter gene.

45 50 Next, the nucleotide sequence of the isolated positive clone is determined. In the case of a cDNA which is proved to code for an unknown protein, the clone of the full length is isolated with the use of the cDNA as a probe and the full nucleotide sequence can be thus identified. All of these operations are carried out by methods which are well known by those skilled in the art. For example, the nucleotide sequence may be identified by the Maxam-Gilbert method or the dideoxy terminator method. On the other hand, the full length may be sequenced in accordance with a method described in Molecular Cloning [Sambrook, J., Fritsch, E.F. and Maniatis, T., published by Cold Spring Harbor Laboratory Press in 1989].

55 The present invention further relates to a novel polypeptide which has been identified using the process of the present invention to construct a cDNA library and a DNA coding for the same. In particular, it relates to:

- (1) a polypeptide consisting of an amino acid sequence represented by SEQ ID No. 1;
- (2) a DNA coding for the polypeptide described in the above (1);

- (3) a DNA having a nucleotide sequence represented by SEQ ID No. 2; and
- (4) a DNA having a nucleotide sequence represented by SEQ ID No. 3.

A polypeptide of Seq. ID No. 1 in substantially purified form will generally comprise the polypeptide in a preparation in which more than 90%, eg. 95%, 98% or 99% of the polypeptide in the preparation is that of the Seq. ID No. 1.

5 A polypeptide homologue of the Seq. ID No. 1 will be generally at least 70%, preferably at least 80 or 90% and more preferably at least 95% homologous to the polypeptide of Seq. ID No. 1 over a region of at least 20, preferably at least 30, for instance 40, 60 or 100 more contiguous amino acids. Such polypeptide homologues will be referred to below as a polypeptide according to the invention.

10 Generally, fragments of Seq. ID No. 1 or its homologues will be at least 10, preferably at least 15, for example 20, 25, 30, 40, 50 or 60 amino acids in length, and are also encompassed by the term "a polypeptide according to the invention" as used herein. Particular fragments of the polypeptides of the invention are fragments of which include amino acid residues numbered 1-70 in Seq ID No. 4 or a homologue thereof.

15 A DNA capable of selectively hybridizing to the DNA of Seq. ID No. 2 or 3 will be generally at least 70%, preferably at least 80 or 90% and more preferably at least 30, for instance 40, 60 or 100 or more contiguous nucleotides. Such DNA will be encompassed by the term "DNA according to the invention".

20 Particular DNA capable of selectively hybridising to the DNA of SEQ ID No. 2 or 3 is the nucleotide residues numbered 139-348 in SEQ ID No. 4 or a fragment thereof.

25 DNA according to the invention may be used to produce a primer, e.g. a PCR primer, a probe e.g. labelled by conventional means using radioactive or non-radioactive labels, or the DNA may be cloned into a vector. Such primers, probes and other fragments of the DNA of Seq. ID No. 2 or 3 will be at least 15, preferably at least 20, for example 25, 30 or 40 nucleotides in length, and are also encompassed by the term "DNA according to the invention" as used herein.

30 DNA according to the invention may be produced recombinantly, synthetically, or by any means available to those of skill in the art.

35 A further embodiment of the invention provides replication and expression vectors comprising DNA according to the invention. The vectors may be, for example, plasmid, virus or phage vectors provided with an origin of replication, optionally a promoter for the expression of the said DNA and optionally a regulator of the promotor. The vector may contain one or more selectable marker genes, for example an ampicillin resistance gene. The vector may be used in vitro, for example for the production of RNA corresponding to the DNA, or used to transform a host cell.

40 A further embodiment of the invention provides host cells transformed or transfected with the vectors for the replication and expression of DNA according to the invention, including the DNA Seq. ID No. 2 or 3 or the open reading frame thereof. The cells will be chosen to be compatible with the vector and may for example be bacterial, yeast, insect or mammalian.

45 DNA according to the invention may also be inserted into the vectors described above in an antisense orientation in order to provide for the production of antisense RNA (DNA). Antisense RNA (DNA) may also be produced by synthetic means. Such antisense RNA (DNA) may be used in a method of controlling the levels of a polypeptide of the invention in a cell.

50 A further embodiment of the present invention provides a method of producing a polypeptide which comprises culturing host cells of the present invention under conditions effective to express a polypeptide of the invention. Preferably, in addition, such a method is carried out under conditions in which the polypeptide of the invention is expressed and then secreted from the host cells.

55 The invention also provides monoclonal or polyclonal antibodies to a polypeptide according to the invention. The invention further provides a process for the production of monoclonal or polyclonal antibodies to the polypeptides of the invention. Monoclonal antibodies may be prepared by conventional hybridoma technology using a polypeptide of the invention or a fragment thereof, as an immunogen. Polyclonal antibodies may also be prepared by conventional means which comprise inoculating a host animal, for example a rat or a rabbit, with a polypeptide of the invention and recovering immune serum.

60 The present invention also provides pharmaceutical compositions containing a polypeptide of the invention, or an antibody thereof, in association with a pharmaceutically acceptable diluent or carrier. The invention also provides a polypeptide according to the invention or an antibody for use in a method of therapy or diagnosis on a human or animal body.

65 The polypeptides of the present invention include not only those having the amino acid sequence represented by the SEQ ID No. 1 but also those with partial deletion thereof (for example, a polypeptide consisting of the mature protein part alone, or consisting of a part of the mature protein essentially required for the expression of the biological activity), those with partial replacement by other amino acid(s) (for example, a polypeptide some of amino acids are replaced by those having similar properties) and those with partial addition or insertion of amino acid(s).

70 It is well known that there are up to six different codons which may code for a single amino acid (for example, one type of codon for Met while six types of codon for Leu). Accordingly, the nucleotide sequence of the DNA can be changed without altering the amino acid sequence of the polypeptide.

The DNA as specified in (2) includes all nucleotide sequences coding for the polypeptide represented by Seq. ID No. 1. Changes in the nucleotide sequence sometimes bring about an increase in the polypeptide productivity.

5 The DNA as specified in (3) is an embodiment of the DNA as specified in (2) and represents a natural sequence.

The DNA as specified in (4) represents a sequence wherein a natural non-translational region is added to the DNA as specified in (3).

10 A signal peptide is a highly hydrophobic region located immediately downstream of the translation initiation region ranging from Met at the 1-position to Ser at the 19-position in the amino acid sequence represented by Seq. ID No. 1. The region essentially responsible for the expression of the biological activity corresponds to the part of the amino acid sequence of the Seq. ID No. 1 lacking of the signal peptide, i.e., the mature protein part. Thus the signal peptide never relates to the activity.

15 The DNA having the nucleotide sequence represented by Seq. ID No. 3 can be prepared in accordance with the process described as the first subject of the present invention.

Once the nucleotide sequences represented by Seq. ID No. 2 and No. 3 are determined, the DNA of the present invention can be chemically synthesized. Alternatively, the DNA of the present invention can be obtained by chemically synthesizing fragments of said nucleotide sequence and hybridizing with the use of the

20 DNA containing said DNA into an appropriate host and then incubating the host.

Examples of methods for obtaining the polypeptide of the present invention include:

- (1) isolation and purification from vital tissues or cultured cells;
- (2) chemical synthesis of peptides; and
- (3) production with the use of gene recombination techniques;

25 described in (3) is preferable.

Examples of the expression system (host-vector system) for producing the polypeptide by using gene recombination techniques include those of bacteria, yeasts, insect cells and mammalian cells.

30 In order to express in *E. coli*, for example, an initiator codon (ATG) is added to the 5'-end of the DNA coding for the mature protein region. The DNA thus obtained is then ligated to the downstream of an appropriate promoter (for example, trp promoter, lac promoter, λ p_L promoter, T7 promoter, etc.) and inserted into a vector capable of functioning in *E. coli* (for example, pBR322, pUC18, pUC19, etc.), thus constructing an expression vector. Next, an *E. coli* strain (for example, *E. coli*, DH1, *E. coli* JM109, *E. coli* HB101, etc.) transformed with this expression vector is incubated in an appropriate medium. Thus the target polypeptide can be obtained from the incubated cells. Alternately, a bacterial signal peptide (for example, a signal peptide of *pelB*) may be used and thus the polypeptide can be secreted into the periplasm. Furthermore, a fusion protein together with other polypeptide can be produced.

35 Expression in mammalian cells can be effected, for example, in the following manner. Namely, a DNA coding for the nucleotide sequence represented by Seq. ID No. 3 is inserted into the downstream of an appropriate promoter (for example, SV40 promoter, LTR promoter, metallothionein promoter, etc.) in an appropriate vector (for example, retrovirus vector, papilloma virus vector, vaccinia virus vector, SV40-series vector, etc.), thus constructing an expression vector. Next, appropriate mammalian cells (for example, monkey COS-7 cells, Chinese hamster CHO cells, mouse L cells, etc.) are transformed with the expression vector obtained above and the transformant is incubated in an appropriate medium. Thus the target polypeptide can be secreted into the culture medium. The polypeptide thus obtained can be isolated and purified by conventional biochemical methods.

40 By using the process for constructing a cDNA library which is the first subject of the present invention, a DNA coding for a signal peptide of a secretory protein or a membrane protein can be efficiently selected and, in its turn, an unknown and useful protein can be efficiently found out. The novel polypeptide which is the second subject of the present invention is produced and secreted from a stroma cell line. Therefore, the polypeptide has biological activities relating to the survival and proliferation of hematopoietic stem cells and the proliferation and differentiation of B cells and myeloid cells, and chemoattractant activity of neutrophil. Accordingly, the polypeptide of the present invention per se is usable as an agent for preventing or treating, for example, anemia or leukopenia, infections, etc.

45 In addition, the above-mentioned polypeptide existing in vivo can be assayed by using a polyclonal antibody or a monoclonal antibody for said polypeptide, which is applicable to studies on the relationship between said polypeptide and diseases or to the diagnosis of diseases and the like. The polyclonal antibody and the monoclonal antibody can be prepared by a conventional method with the use of said polypeptide or a fragment thereof as an antigen.

5 The DNA according to the present invention serves as an important and essential template in the production of the polypeptide of the present invention which is expected to be highly useful. Further, the DNA of the present invention is applicable to the diagnosis and treatment of hereditary diseases, i.e., gene therapy, and therapy with ceasing the expression of the polypeptide by using antisense DNA (RNA). Furthermore, a genomic DNA can be isolated by using the DNA of the present invention as a probe. Similarly, a human gene for a related polypeptide being highly homologous with the DNA of the present invention and a gene of an organism other than human for a polypeptide being high homologous with the polypeptide of the present invention can be isolated.

10 Examples

The following Examples and Reference Example are illustrated, but not limit the present invention.

Reference Example 1

15 The following Examples and References describe the construction and expression of plasmid pcDL-SR α -h-G-CSF-hTac (pSGT) and plasmid pcDL-SR α -hRAR α -hTac (pSRT). Construction and expression of plasmid pcDL-SR α -h-G-CSF-hTac (pSGT) and plasmid pcDL-SR α -hRAR α -hTac (pSRT) wherein a cDNA coding for a fusion protein of hG-CSF (human granulocyte colony stimulating factor) having a signal peptide) or hRAR α (human retinoic acid receptor α , a member of the nuclear receptor superfamily) is fused with hTac (human IL-2 receptor α , used as a reporter gene) under the control of the SV40 early promoter. The plasmid pSGT was constructed by fusing the SV40 early promoter with the hG-CSF cDNA, and the plasmid pSRT was constructed by fusing the SV40 early promoter with the hRAR α cDNA.

Reference Example 1
Construction and expression of plasmid pcDL-SR α -h-G-CSF-hTac (pSGT) and plasmid hTac (pSRT)
A plasmid, wherein a cDNA coding for a fusion protein of hG-CSF (human granulocyte colony stimulating factor, a typical example of a protein having a signal peptide) or hRAR α (human retinoic acid receptor α , a typical example of a protein having no signal peptide), with hTac (human IL-2 receptor α , used as a reporter gene) was integrated into an eucaryotic cell expression plasmid vector pcDL-SR α having an SR α promoter [described in Mol. Cell. Biol., 8, 466 (1988), provided by Dr. Yutaka Takebe, National Institute of Health], was constructed. After transformation, the expression of the reporter protein on the membrane was examined.
(1) By employing a plasmid pSP72-hG-CSF, wherein hG-CSF cDNA had been integrated into the EcoRI site of a plasmid pSP72 (purchased from Promega), as a template and using an SP6 promoter primer (purchased from Takara Shuzo Co., Ltd.) and an hG-CSF specific primer having an SacI site added thereto,

5' GGAGATATC GAGCTCCTGGGGTGGCACAG 3'
 EcoRV SacI hG-CSF cDNA antisense

0

35

40

PCR was performed 25 cycles (at 95 °C for one minute, at 48 °C for two minutes and at 72 °C for two minutes). The amplified DNA fragment was digested with SacI-EcoRI and once subcloned into a plasmid pBlue script SK(+) (pPBS). After digesting with SacI-EcoRI again, an EcoRI-Sacl fragment of hG-CSF was obtained. On the other hand, a plasmid pBS-hTac, wherein hTac cDNA had been integrated into the HindIII site of pPBS, was digested with SacI-KpnI to thereby give an SacI-KpnI fragment of hTac cDNA with the deletion of the signal sequence. These fragments were integrated into the EcoRI-KpnI site of pcDL-SR α with the deletion of stuffer (Fig. 2) to thereby give a plasmid pcDL-SR α -hG-CSF-hTac (pSGT) (Figs. 3, 4 and 5).

45

50

55

60

65

70

75

80

85

90

95

PCR was performed 25 cycles (at 95 °C for one minute, at 48 °C for two minutes and at 72 °C for two minutes). The amplified DNA fragment was digested with SacI-EcoRI and once subcloned into a plasmid pBlue script SK(+) (pPBS). After digesting with SacI-EcoRI again, an EcoRI-Sacl fragment of hG-CSF was obtained. On the other hand, a plasmid pBS-hTac, wherein hTac cDNA had been integrated into the HindIII site of pPBS, was digested with SacI-KpnI to thereby give an SacI-KpnI fragment of hTac cDNA with the deletion of the signal sequence. These fragments were integrated into the EcoRI-KpnI site of pcDL-SR α with the deletion of stuffer (Fig. 2) to thereby give a plasmid pcDL-SR α -hG-CSF-hTac (pSGT) (Figs. 3, 4 and 5).

95

100

105

110

115

120

125

130

135

140

145

150

155

160

165

170

175

180

185

190

195

200

205

210

215

220

225

230

235

240

245

250

255

260

265

270

275

280

285

290

295

300

305

310

315

320

325

330

335

340

345

350

355

360

365

370

375

380

385

390

395

400

405

410

415

420

425

430

435

440

445

450

455

460

465

470

475

480

485

490

495

500

505

510

515

520

525

530

535

540

545

550

555

560

565

570

575

580

585

590

595

600

605

610

615

620

625

630

635

640

645

650

655

660

665

670

675

680

685

690

695

700

705

710

715

720

725

730

735

740

745

750

755

760

765

770

775

780

785

790

795

800

805

810

815

820

825

830

835

840

845

850

855

860

865

870

875

880

885

890

895

900

905

910

915

920

925

930

935

940

945

950

955

960

965

970

975

980

985

990

995

1000

1005

1010

1015

1020

1025

1030

1035

1040

1045

1050

1055

1060

1065

1070

1075

1080

1085

1090

1095

1100

1105

1110

1115

1120

1125

1130

1135

1140

1145

1150

1155

1160

1165

1170

1175

1180

1185

1190

1195

1200

1205

1210

1215

1220

1225

1230

1235

1240

1245

1250

1255

1260

1265

1270

1275

1280

1285

1290

1295

1300

1305

1310

1315

1320

1325

1330

1335

1340

1345

1350

1355

1360

1365

1370

1375

1380

1385

1390

1395

1400

1405

1410

1415

1420

1425

1430

1435

1440

1445

1450

1455

1460

1465

1470

1475

1480

1485

1490

1495

1500

1505

1510

1515

1520

1525

1530

1535

1540

1545

1550

1555

1560

1565

1570

1575

1580

1585

1590

1595

1600

1605

1610

1615

1620

1625

1630

1635

1640

1645

1650

1655

1660

1665

1670

1675

1680

1685

1690

1695

1700

1705

1710

1715

1720

1725

1730

1735

1740

1745

1750

1755

1760

1765

1770

1775

1780

1785

1790

1795

1800

1805

1810

1815

1820

1825

1830

1835

1840

1845

1850

1855

1860

1865

1870

1875

1880

1885

1890

1895

1900

1905

1910

1915

1920

1925

1930

1935

1940

1945

1950

1955

1960

1965

1970

1975

1980

1985

1990

1995

2000

2005

2010

2015

2020

2025

2030

2035

2040

2045

2050

2055

2060

2065

2070

2075

2080

2085

2090

2095

2100

2105

2110

2115

2120

2125

2130

2135

2140

2145

2150

2155

2160

2165

2170

2175

2180

2185

2190

2195

2200

2205

2210

2215

2220

2225

2230

2235

2240

2245

2250

2255

2260

2265

2270

2275

2280

2285

2290

2295

2300

2305

2310

2315

2320

2325

2330

2335

2340

2345

2350

2355

2360

2365

2370

2375

2380

2385

2390

2395

2400

2405

2410

2415

2420

2425

2430

2435

2440

2445

2450

2455

2460

2465

2470

2475

2480

2485

2490

2495

2500

2505

2510

2515

2520

2525

2530

2535

2540

2545

2550

2555

2560

2565

2570

2575

2580

2585

2590

2595

2600

2605

2610

2615

2620

2625

2630

2635

2640

2645

2650

2655

2660

2665

2670

2675

2680

2685

2690

2695

2700

2705

2710

2715

2720

2725

2730

2735

2740

2745

2750

2755

2760

2765

2770

2775

2780

2785

2790

2795

2800

2805

2810

2815

2820

2825

2830

2835

2840

2845

2850

2855

2860

2865

2870

2875

2880

2885

2890

2895

2900

2905

2910

2915

2920

2925

2930

2935

2940

2945

2950

2955

2960

2965

2970

2975

2980

2985

2990

2995

3000

3005

3010

3015

3020

3025

3030

3035

3040

3045

3050

3055

3060

3065

3070

3075

3080

3085

3090

3095

3100

3105

3110

3115

3120

3125

3130

3135

3140

3145

3150

3155

3160

3165

3170

3175

3180

3185

3190

3195

3200

3205

3210

3215

3220

3225

3230

3235

3240

3245

3250

3255

3260

3265

3270

3275

3280

3285

3290

3295

3300

3305

3310

3315

3320

3325

3330

3335

3340

3345

3350

3355

3360

3365

3370

3375

3380

3385

3390

3395

3400

3405

3410

3415

3420

3425

3430

3435

3440

3445

3450

3455

3460

3465

3470

3475

3480

3485

3490

3495

3500

3505

3510

3515

3520

3525

3530

3535

3540

3545

3550

3555

3560

3565

3570

3575

3580

3585

3590

3595

3600

3605

3610

3615

3620

3625

3630

3635

3640

3645

3650

3655

3660

3665

3670

3675

3680

3685

3690

3695

3700

3705

3710

3715

3720

3725

3730

3735

3740

3745

3750

3755

3760

3765

3770

3775

3780

3785

3790

3795

3800

3805

3810

3815

3820

3825

3830

3835

3840

3845

3850

3855

3860

3865

3870

3875

3880

3885

3890

3895

3900

3905

3910

3915

3920

3925

3930

3935

3940

3945

3950

3955

3960

3965

3970

3975

3980

3985

3990

3995

4000

4005

4010

4015

4020

4025

4030

4035

4040

4045

4050

4055

4060

4065

4070

4075

4080

4085

4090

4095

4100

4105

4110

4115

4120

4125

4130

4135

4140

4145

4150

4155

4160

4165

4170

4175

4180

4185

4190

4195

4200

4205

4210

4215

4220

4225

4230

4235

4240

4245

4250

4255

4260

4265

4270

4275

4280

4285

4290

4295

4300

4305

4310

4315

4320

4325

4330

4335

4340

4345

4350

4355

4360

4365

4370

4375

4380

4385

4390

4395

4400

4405

4410

4415

4420

4425

4430

4435

4440

4445

4450

4455

4460

4465

4470

4475

4480

4485

4490

4495

4500

4505

4510

4515

4520

4525

4530

4535

4540

4545

4550

4555

4560

4565

4570

4575

4580

4585

4590

4595

4600

4605

4610

4615

4620

4625

4630

4635

4640

4645

4650

4655

4660

4665

4670

4675

4680

4685

4690

4695

4700

4705

4710

4715

4720

4725

4730

4735

4740

4745

4750

4755

4760

4765

4770

4775

4780

4785

4790

4795

4800

4805

4810

4815

4820

4825

4830

4835

4840

4845

4850

4855

4860

4865

4870

4875

4880

4885

4890

4895

4900

4905

4910

4915

4920

4925

4930

4935

4940

4945

4950

4955

4960

4965

4970

4975

4980

4985

4990

4995

5000

5005

5010

5015

5020

5025

5030

5035

5040

5045

5050

5055

5060

5065

5070

5075

5080

5085

5090

5095

5100

5105

5110

5115

5120

5125

5130

5135

5140

5145

5150

5155

5160

5165

5170

5175

5180

5185

5190

5195

5200

5205

5210

5215

5220

5225

5230

5235

5240

5245

5250

5255

5260

5265

5270

5275

5280

5285

5290

5295

5300

5305

5310

5315

5320

5325

5330

5335

5340

5345

5350

5355

5360

5365

5370

5375

5380

5385

5390

5395

5400

5405

5410

5415

5420

5425

5430

5435

5440

5445

5450

5455

5460

5465

5470

5475

5480

5485

5490

5495

5500

5505

5510

5515

5520

5525

5530

5535

5540

5545

5550

5555

5560

5565

5570

5575

5580

5585

5590

5595

5600

5605

5610

5615

5620

5625

5630

5635

5640

5645

5650

5655

5660

5665

5670

5675

5680

5685

5690

5695

5700

5705

5710

5715

5720

5725

5730

5735

5740

5745

5750

5755

5760

5765

5770

5775

5780

5785

5790

5795

5800

5805

5810

5815

5820

5825

5830

5835

5840

5845

5850

5855

5860

5865

5870

5875

5880

5885

5890

5895

5900

5905

5910

5915

5920

5925

5930

5935

5940

5945

5950

5955

5960

5965

5970

5975

5980

5985

5990

5995

6000

6005

6010

6015

6020

6025

6030

6035

6040

6045

6050

6055

6060

6065

6070

6075

6080

6085

6090

6095

6100

6105

6110

6115

6120

6125

6130

6135

6140

6145

6150

6155

6160

6165

6170

6175

6180

6185

6190

6195

6200

6205

6210

6215

6220

6225

6230

6235

6240

6245

6250

6255

6260

6265

6270

6275

6280

6285

6290

6295

6300

6305

6310

6315

6320

6325

6330

6335

6340

6345

6350

6355

6360

6365

6370

6375

6380

6385

6390

6395

6400

6405

6410

6415

6420

6425

6430

6435

6440

6445

6450

6455

6460

6465

6470

6475

6480

6485

6490

6495

6500

6505

6510

6515

6520

6525

6530

6535

6540

6545

6550

6555

6560

6565

6570

6575

6580

6585

6590

6595

6600

6605

6610

6615

6620

6625

6630

6635

6640

6645

6650

6655

6660

6665

6670

6675

6680

6685

6690

6695

6700

6705

6710

6715

6720

6725

6730

<p

Next, by employing a plasmid pGEM3-hRAR α , wherein hRAR α cDNA had been integrated into the EcoRI site of a plasmid pGEM3, as a template and using an SP6 promoter primer and an hRAR α specific primer having an SacI site added thereto,

45 ECORI primer having an SacI site and
 5' GGAGATATC GAGCTCAATGGTGGCTGGGGATG 3'
 EcoRV SacI
 hRAR α cDNA antisense

50 PCR was performed. Subsequently, the procedure employed in the above-mentioned case of G-CSF was re-
55 peated to thereby give a plasmid pcDL-SR α -hRAR α -hTac (pSRT) (Figs. 6, 4 and 5).
(2) The pSGT or pSRT obtained in the above (1) was transfected into COS-7 cells by the DEAE-dextran
method [described in detail in Current Protocol in Molecular Biology, chapter 9.2.1.]. After 48 hours, the
cells were harvested from a dish and incubated together with mouse anti-Tac IgG antibody for 20 minutes
on ice. After eliminating free antibodies, the mixture was incubated together with goat anti-mouse IgG for 20 minutes

tibody labeled with fluorescein isothiocyanate (FITC) for 20 minutes on ice. After eliminating free antibodies again, the expression of a fusion protein G-CSF-Tac or RAR α -Tac on the membrane was examined with a fluorescence activated cell sorter (Model FACS Can, manufactured by BECTON DICKINSON, hereinafter referred to simply as FACS). Figs. 7 and 8 show the results of the judgement.

As shown in Fig. 7, G-CSF-Tac was expressed on the membrane as well as Tac. As shown in Fig. 8, on the other hand, RAR α -Tac was not detected on the membrane but remained within the cells. These results indicate that when a cDNA fragment containing a signal peptide is ligated to the upstream of a reporter gene, said reporter protein is expressed on the cell membrane, while when a cDNA fragment having no signal peptide is ligated, the reporter protein is not expressed.

Example 1

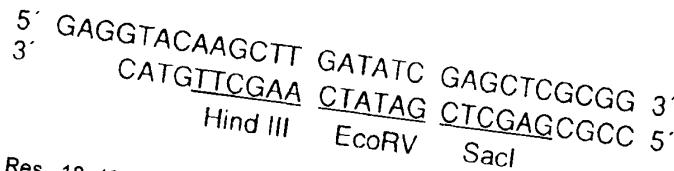
Construction of cDNA library having selectivity for signal peptides (Figs. 9 and 10)

15 Total RNA was extracted from a mouse stroma cell line ST2 [cells supporting the survival and proliferation of hematopoietic stem cells and the proliferation and differentiation of B cells and myeloid cells; refer to EMBO J., 7, 1337 (1988)] by the acid guanidine-phenol-chloroform (AGPC) method [described in detail in "Saibo Kogaku Jikken Protokoru (Protocol in Cellular Engineering Experiments)", published by Shujun-sha, 28 - 31].

20 Then poly A-RNA was purified by using oligo (dT)-latex (Oligotex-dT30®, marketed from Takara Shuzo Co., Ltd.). By using a random hexamer as a primer, a single-stranded cDNA was synthesized with reverse transcriptase and dT was added to the 3'-end thereof with the use of terminal deoxynucleotidyl transferase. A 17 mer dA ligated to a restriction enzyme site containing EcoRI



30 was annealed and a double-stranded cDNA was synthesized by using the same as a primer. Then the cDNA was fragmentated by ultrasonication so as to give an average length of 300 bp and the cDNAs of 200 to 500 bp were fractionated by agarose gel electrophoresis. After blunting the ends with T4DNA polymerase, a lone linker containing an SacI site



40 [see Nucleic Acids Res., 18, 4293 (1990)] was ligated and cDNAs of 200 to 500 bp were fractionated again by agarose gel electrophoresis. By using a primer (NLC) containing an EcoRI site

NLC

45 and another primer (LLHES) containing an SacI site

LLHES



PCR was performed 25 cycles (at 94 °C for one minute, at 50 °C for two minutes and at 72 °C for two minutes). The amplified cDNA was digested with SacI and EcoRI and cDNAs of 250 to 500 bp were fractionated by agarose gel electrophoresis. The cDNA was ligated to a plasmid obtained by digesting pSRT (prepared in Reference Example 1) with SacI and EcoRI by using T4 DNA ligase. After transformation of an *E. coli* DH5 α strain, a cDNA library having a selectivity for signal peptides was obtained.

Example 2

Screening and analysis of cDNA coding for signal peptide

About 1,200 colonies in the library obtained in Example 1 were divided into 24 pools (about 50 colonies/pool). Plasmids of each pools were isolated by the miniprep method and transfected into COS-7 cells by the DEAE-dextran method. After 48 to 72 hours, cell surface-staining for Tac of the transfectant was performed in the same manner as described in Reference Example 1 and 6 positive pools were selected under a fluorescent microscope. Colonies of one pool from among the 6 positive pools were further divided and the same procedure as described above was repeated until a single clone was obtained. Thus a positive clone (pS-TT3) was obtained. Subsequently, by using two synthetic primers

5' TTTACTTCTAGGCCTGTACG 3'

(20 bases upstream from EcoRI cloning site, for sense)
and

5' CCATGGCTTGAATGTGGCG 3'

(20 bases downstream from SacI cloning site, for antisense)

which were specific for the pcDL-SR α -Tac vector, the nucleotide sequence of the TT3 insert was determined. An open reading frame following the Tac cDNA with the deletion of the signal sequence in-frame was searched and converted into the deduced amino acid sequence. After performing a hydrophobicity profile, it was confirmed that a hydrophobic region characteristic to a signal peptide was contained therein (Fig. 11). Further, the homology with data base on DNA and amino acid levels was examined. As a result, it has been found out that TT3 clone codes for an unknown protein.

Example 3

Screening of cDNA with the full length and determination of nucleotide sequence

A cDNA library was constructed by using Super Script[®] Ramda System (marketed from BRL). Next, pS-TT3 was digested with SacI and EcoRI and a TT3 cDNA fragment was prepared by agarose gel electrophoresis. The library was screened by using an oligo-labeled TT3 cDNA fragment as a probe and thus a number of positive clones were obtained. Among these clones, a TT3-1-6 clone showing the longest insert was selected and an Sall-NotI fragment excised from a λ gt22A vector was subcloned into PBS to thereby give a plasmid pBS-TT316. By using a T7 primer, the nucleotide sequence of 300 bp in the 5'-terminal of TT3-1-6 cDNA was determined. Thus it was confirmed that the sequence identical with TT3 of the probe existed in the most 5'-end of TT3-1-6.

Next, a number of pBS-TT316 variant plasmids lacking of the 5-end or the 3'-end of TT3-1-6 cDNA were constructed by using an ExoIII/Mung Bean Deletion Kit (manufactured by Stratagene). By using these variant plasmids, the nucleotide sequence of the full length of the cDNA was determined (sequence No. 3). From the full length cDNA sequence data, an open reading frame was determined and further translated into an amino acid sequence. Thus the sequence represented by Seq. ID No. 1 was obtained. The amino acid sequence at the 30- to 40-positions in the N-terminal of the amino acid sequence thus obtained were compared with known signal peptides. Thus the signal peptide part of this polypeptide was deduced to thereby give the sequence represented by Seq. ID No. 4.

SEQUENCE LISTING

5 GENERAL INFORMATION:

APPLICANT:

10 NAME: Ono Pharmaceutical Co., Ltd.
 STREET: 1-5 Doshomachi 2-chome
 CITY: Osaka

15 COUNTRY: Japan
 POSTAL CODE (ZIP): 541

20 TITLE OF THE INVENTION:

25 Process for constructing cDNA library, and novel polypeptide and
 DNA coding for the same

NUMBER OF SEQUENCES: 4

30 SEQUENCE NO: 1

LENGTH: 89 amino acids

35 TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

40 SEQUENCE: Seq. ID No. 1

Met Asp Ala Lys Val Val Ala Val Leu Ala Leu Val Leu Ala Ala Leu
 1 5 10 15
 Cys Ile Ser Asp Gly Lys Pro Val Ser Leu Ser Tyr Arg Cys Pro Cys
 20 25 30
 Arg Phe Glu Ser His Ile Ala Arg Ala Asn Val Lys His Leu Lys
 35 40 45
 Ile Leu Asn Thr Pro Asn Cys Ala Leu Gln Ile Val Ala Arg Leu Lys
 50 55 60
 Asn Asn Asn Arg Gln Val Cys Ile Asp Pro Lys Leu Lys Trp Ile Gln
 65 70 75 80
 Glu Tyr Leu Glu Lys Ala Leu Asn Lys

SEQUENCE NO: 2

LENGTH: 267 base pairs

TYPE: nucleic acid

5 STRANDEDNESS: single

TOPOLOGY: linear

10 MOLECULE TYPE: cDNA to mRNA

SEQUENCE: Seq. ID No. 2

15	ATGGACGCCA AGGTCGTCGC CGTGCTGGCC CTGGTGCTGG CCGCGCTCTG CATCAGTGAC	60
	GGTAAACCAAG TCAGCCTGAG CTACCGATGC CCCTGCCGGT TCTTCGAGAG CCACATGCC	120
	AGAGCCAACG TCAAGCATCT GAAAATCCTC AACACTCAA ACTGTGCCCT TCAGATTGTT	180
20	GCACGGCTGA AGAACAAACAA CAGACAAGTG TGCATTGACC CGAAATTAAA GTGGATCCAA	240
	GAGTACCTGG AGAAAGCTTT AAACAAG	267

25 SEQUENCE NO: 3

LENGTH: 1797 base pairs

TYPE: nucleic acid

30 STRANDEDNESS: single

TOPOLOGY: linear

35 MOLECULE TYPE: cDNA to mRNA

SEQUENCE: Seq. ID No. 3

40	GACCACTTTC CCTCTCGGTC CACCTCGGTG TCCTCTTGCT GTCCAGCTCT GCAGCCTCCG	60
	GCGCGCCCTC CCGCCCACGC CATGGACGCC AAGGTCGTCG CCGTGCTGGC CCTGGTGCTG	120
	GCCGCGCTCT GCATCAGTGA CGGTAAACCA GTCAAGCCTGA GCTACCGATG CCCCTGCCGG	180
45	TTCTTCGAGA GCCACATCGC CAGAGCAAAC GTCAAGCATC TGAAAATCCT CAACACTCCA	240
	AACTGTGCC C TTCAAGATTGT TGCACGGCTG AAGAACAAACA ACAGACAAGT GTGCATTGAC	300
	CCGAAATTAA AGTGGATCCA AGAGTACCTG GAGAAAGCTT TAAACAAGTA AGCACAAACAG	360
	CCCAAAGGAC TTTCAGTAG ACCCCCAGG AAGGCTGACA TCCGTGGGAG ATGCAAGGGC	420
50	AGTGGTGGGG AGGAGGGCCT GAACCCCTGGC CAGGATGGCC GGCGGGACAG CACTGACTGG	480
	GGTCATGCTA AGGTTTGCCA GCATAAAGAC ACTCCGCCAT AGCATATGGT ACGATATTGC	540
	AGCTTATATT CATCCCTGCC CTCGCCCGTG CACAATGGAG CTTTTATAAC TGGGTTTTT	600
55	CTAAGGAATT GTATTACCT ACCAGTTAG CTTCATCCCC ATTCTCCCTCA TCCTCATCTT	660

5 CATTTTAAAA AGCAGTGATT ACTTCAAGGG CTGTATTCAAG TTTGCTTGG AGCTTCTCTT 720
 TGCCCTGGGG CCTCTGGCA CAGTTATAGA CGGTGGCTT GCAGGGAGCC CTAGAGAGAA 780
 ACCTTCCACC AGAGCAGAGT CCGAGGAACG CTGCAGGGCT TGTCCTGCAG GGGGCGCTCC 840
 TCGACAGATG CCTTGTCTG AGTCAACACA AGATCCGGCA GAGGGAGGCT CCTTTATCCA 900
 GTTCAGTGCCTC AGGGTCGGGA AGCTTCCTT AGAAGTGATC CCTGAAGCTG TGCTCAGAGA 960
 CCCTTCTCTA GCCGTTCTG CTCTCTGCTT GCCTCCAAAC GCATGCTTCATC TCTGACTTCC 1020
 GCTTCTCACC TCTGTAGCCT GACGGACAA TGCTGCAATG GAAGGGAGGA GAGTGATGTG 1080
 GGGTGCCCCC TCCCTCTCTT CCCTTGCTT TCCTCTCACT TGGGCCCTT GTGAGATTT 1140
 TCTTTGGCCT CCTGTAGAAT GGAGGCCAGAC CATCCTGGAT AATGTGAGAA CATGCCTAGA 1200
 10 TTTACCCACA AAACACAAGT CTGAGAATTA ATCATAAACG GAAGTTAAA TGAGGATTTG 1260
 GACCTTGGTA ATTGTCCCTG AGTCCTATAT ATTTCAACAG TGGCTCTATG GGCTCTGATC 1320
 GAATATCAGT GATGAAAATA ATAATAATAA TAATAATAAC GAATAAGCCA GAATCTGCC 1380
 ATGAAGCCAC AGTGGGGATT CTGGGTTCCA ATCAGAAATG GAGACAAGAT AAAACTTGCA 1440
 15 TACATTCTTA TGATCACAGA CGGCCCTGGT GGTTTTGGT AACTATTTAC AAGGCATTT 1500
 TTACATATA TTTTGTCGA CTTTTATGT TTCTTTGGAA GACAAATGTA TTTCAGAATA 1560
 TATTTGAGT CAATTCTAT ATTTGAAGTG GAGCCATAGT AATGCCAGTA GATATCTCTA 1620
 TGATCTTGAG CTACTGGCAA CTTGTAAAGA AATATATATG ACATATAAAAT GTATTGTAGC 1680
 20 TTCCGGTGT CAGCCACGGT GTATTTTCC ACTTGGAAATG AAATTGTATC AACTGTGACA 1740
 TTATATGCAC TAGCAATAAA ATGCTAATTG TTTCATGCTG TAAAAAAA AAAAAAA 1797
 25
 30

SEQUENCE NO: 4

LENGTH: 1797 base pairs

35 TYPE: nucleic acid

STRANDEDNESS: single

40 TOPOLOGY: linear

MOLECULE TYPE: cDNA to mRNA

ORIGINAL SOURCE

45 ORGANISM: Mouse

CELL LINE: ST2

FEATURE

50 NAME/KEY: CDS

LOCATION: 82 .. 351

55 IDENTIFICATION METHOD: P

NAME/KEY: sig peptide

LOCATION: 82 .. 138

IDENTIFICATION METHOD: S

5

NAME/KEY: mat peptide

10

LOCATION: 139 .. 348

IDENTIFICATION METHOD: S

15

SEQUENCE: Seq. ID No. 4

GACCACTTTC CCTCTCGTC CACCTCGGTG TCCTCTTGCT GTCCAGCTCT GCAGCCTCCG 60
 GCGCGCCCTC CCGCCCACGC C ATG GAC GCC AAG GTC GTC GCC GTG CTG GCC 111
 Met Asp Ala Lys Val Val Ala Val Leu Ala -10
 CTG GTG CTG GCC GCG CTC TGC ATC AGT GAC GGT AAA CCA GTC AGC CTG 159
 Leu Val Leu Ala Ala Leu Cys Ile Ser Asp Gly Lys Pro Val Ser Leu
 AGC TAC CGA TGC CCC TGC CGG TTC TTC GAG AGC CAC ATC GCC AGA GCC 207
 Ser Tyr Arg Cys Pro Cys Arg Phe Phe Glu Ser His Ile Ala Arg Ala
 AAC GTC AAG CAT CTG AAA ATC CTC AAC ACT CCA AAC TGT GCC CTT CAG 255
 Asn Val Lys His Leu Lys Ile Leu Asn Thr Pro Asn Cys Ala Leu Gln
 ATT GTT GCA CGG CTG AAG AAC AAC AGA CAA GTG TGC ATT GAC CCG 303
 Ile Val Ala Arg Leu Lys Asn Asn Arg Gln Val Cys Ile Asp Pro
 AAA TTA AAG TGG ATC CAA GAG TAC CTG GAG AAA GCT TTA AAC AAG TAA 351
 Lys Ile Lys Trp Ile Gln Glu Tyr Leu Glu Lys Ala Leu Asn Lys
 GCACAAACAGC CCAAAGGACT TTCCAGTACA CCCCCGAGGA AGGCTGACAT CCGTGGGAGA 411
 TGCAAGGGCA GTGGTGGGGA GGAGGGCCTG AACCTGGCC AGGATGGCCG GCGGGACAGC 471
 ACTGACTGGG GTCATGCTAA GGTTGCCAG CATAAAGACA CTCCGCCATA GCATATGGTA 531
 CGATATTGCA GCTTATATTTC ATCCCTGCC TCGCCCGTGC ACAATGGAGC TTTTATAACT 591
 GGGGTTTTTC TAAGGAATTG TATTACCTA ACCAGTTAGC TTCAATCCCCA TTCTCCTCAT 651
 CCTCATCTTC ATTTTAAAAA GCAGTGATTA CTCAGGGC TGTATTCACT TTGCTTTGGA 711
 GCTTCTCTT GCCCTGGGGC CTCTGGGCAC AGTTATAGAC GGTGGCTTG CAGGGAGCCC 771
 TAGAGAGAAA CCTTCCACCA GAGCAGAGTC CGAGGAACGC TGCAGGGCTT GTCTGCAGG 831
 GGGCGCTCCT CGACAGATGC CTTGTCTGA GTCAACACAA GATCCGGCAG AGGGAGGCTC 891

5 CTTTATCCAG TTCAGTGCCTA GGGTCGGGAA GCTTCCTTA GAAGTGATCC CTGAAGCTGT 951
 GCTCAGAGAC CCTTCCTAG CCGTCCTGC TCTCTGCTTG CCTCCAAACG CATGCTTCAT 1011
 CTGACTTCCG CTTCTCACCT CTGTAGCCTG ACGGACCAAT GCTGCAATGG AAGGGAGGAG 1071
 AGTGATGTGG GGTGCCCCCT CCCTCTCTTC CCTTGCTTT CCTCTCACTT GGGCCCTTG 1131
 TGAGATTTTT CTTTGGCCTC CTGTAGAATG GAGCCAGACC ATCCTGGATA ATGTGAGAAC 1191
 10 ATGCCTAGAT TTACCCACAA AACACAAGTC TGAGAAATTAA TCATAAACGG AAGTTAAAT 1251
 GAGGATTGG ACCTTGGTAA TTGTCCCTGA GTCCTATATA TTCAACAGT GGCTCTATGG 1311
 GCTCTGATCG AATATCAGTG ATGAAAATAA TAATAATAAT AATAATAACG AATAAGCCAG 1371
 15 AATCTGCCA TGAAGCCACA GTGGGGATTG TGGGTCCTAA TCAGAAATGG AGACAAGATA 1431
 AGGCATTTTT TTACATATAT TTTTGTGCAC TTTTATGTT TCTTGGAAAG ACAAAATGTAT 1491
 TTCAGAATAT ATTTGTAGTC AATTCAATATA TTTGAAGTGG AGCCATAGTA ATGCCAGTAG 1551
 20 ATATCTCTAT GATCTGAGC TACTGGCAC TTGTAAGAA ATATATATGA CATATAAAATG 1611
 TATTGTAGCT TTCCGGTGTG AGCCACGGTG TATTTTCCA CTTGGAATGA AATTGTATCA 1671
 ACTGTGACAT TATATGCACT AGCAATAAAA TGCTAATTGT TTCATGCTGT AAAAAAAAAAA 1731
 AAAAAAA 1791
 25

1797

Claims

30 1. A process for constructing a cDNA library comprising:
 (1) synthesizing single-stranded DNA complementary to mRNA isolated from subject cells using a random polymer and ligating oligo dT to the 3'-end of the single-stranded DNA;
 (2) converting the single-stranded DNA obtained in (1) to a double-stranded DNA using as primer a poly A oligomer ligated to a first restriction enzyme (enzyme I) site;
 (3) fragmenting the double-stranded DNA obtained in (2), fractionating the fragments obtained by size, ligating linker containing a second restriction enzyme (enzyme II) site, differing from the enzyme I, thereto and fractionating;
 (4) amplifying fragments containing the sites of restriction enzymes I and II by a polymerase chain reaction using a first primer containing the enzyme I site and a second primer containing the enzyme II site, digesting the cDNA thus amplified with the restriction enzymes, enzyme I and enzyme II and fractionating; and
 (5) ligating the cDNA fragment thus obtained upstream of a gene coding for a secretory protein or membrane protein having deleted therefrom the sequence coding for a signal peptide, integrating the ligated DNA into a eucaryotic cell expression plasmid vector, and transforming the vector into a host.

35 2. A process according to claim 1, wherein EcoRI is used as the enzyme I, SacI is used as the enzyme II, and human IL-2 receptor α gene is used as a known secretory protein or membrane protein gene with the deletion of signal peptide.

40 3. A polypeptide having the amino acid sequence shown in SEQ ID No. 1 in substantially purified form, a homologue thereof or a fragment of the sequence or homologue of a fragment.

45 4. A polypeptide according to claim 1 having the amino acid sequence shown in SEQ ID No. 1.

50 5. A polypeptide comprising a polypeptide according to claim 3 or 4, having the amino acid sequence of residues 1 to 70 shown in SEQ ID No. 4 in substantially purified form, a homologue thereof, or a fragment of the sequence or homologue of a fragment.

6. A polypeptide according to claim 5 having the amino acid sequence shown of residues 1 to 70 shown in SEQ ID No. 4.
7. DNA encoding a polypeptide according to any one of the preceding claims.
- 5 8. DNA according to claim 7 having the nucleotide sequence shown in SEQ ID No. 2 or a fragment thereof capable of selectively hybridizing to SEQ ID No. 2.
9. DNA according to claim 7 having the nucleotide sequence shown in SEQ ID No. 3 or a fragment thereof capable of selectively hybridizing to SEQ ID No. 3.
- 10 10. DNA according to claim 7 having the nucleotide sequence of residues 139 to 348 shown in SEQ ID No. 4, or a fragment thereof capable of selectively hybridizing to the nucleotide sequence of residues 139 to 348 shown in SEQ ID No. 4.
- 15 11. DNA according to claim 10 having the nucleotide sequence of residues 139 to 348 shown in SEQ ID No. 4.
12. A replication and expression vector comprising DNA according to any one of claims 7 to 11.
- 20 13. Host cells transformed or transfected with a replication and expression vector according to claim 12.
14. A method of producing a polypeptide which comprises culturing host cells according to claim 13 under conditions effective to express a polypeptide according to any one of claims 3 to 6.
- 25 15. A monoclonal or polyclonal antibody to a polypeptide according to any one of claims 3 to 6.
16. A pharmaceutical composition containing a polypeptide according to any one of claims 3 to 6 or an antibody according to claim 15 in association with a pharmaceutically acceptable diluent or carrier.

30

35

40

45

50

55

Figure 1

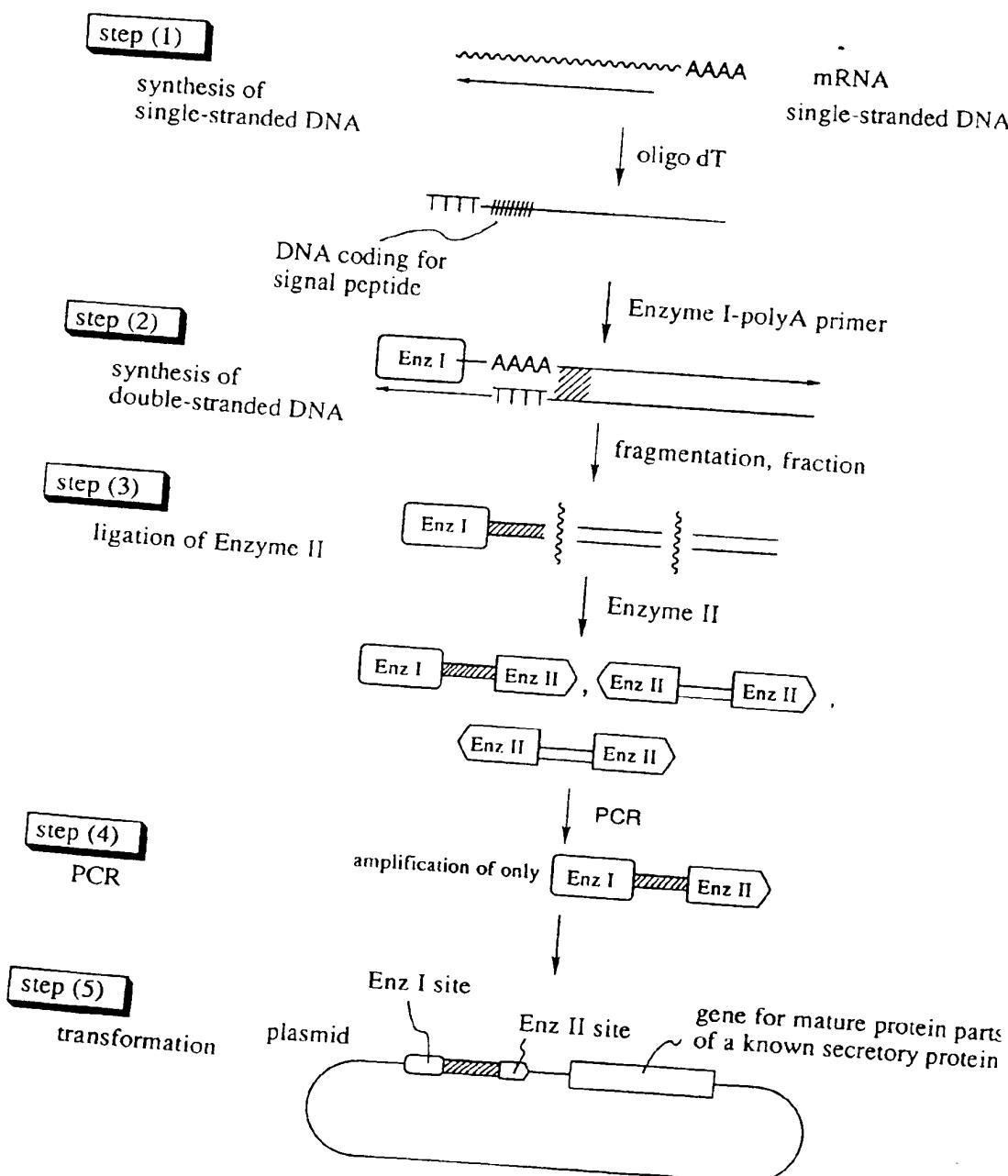


Figure 2

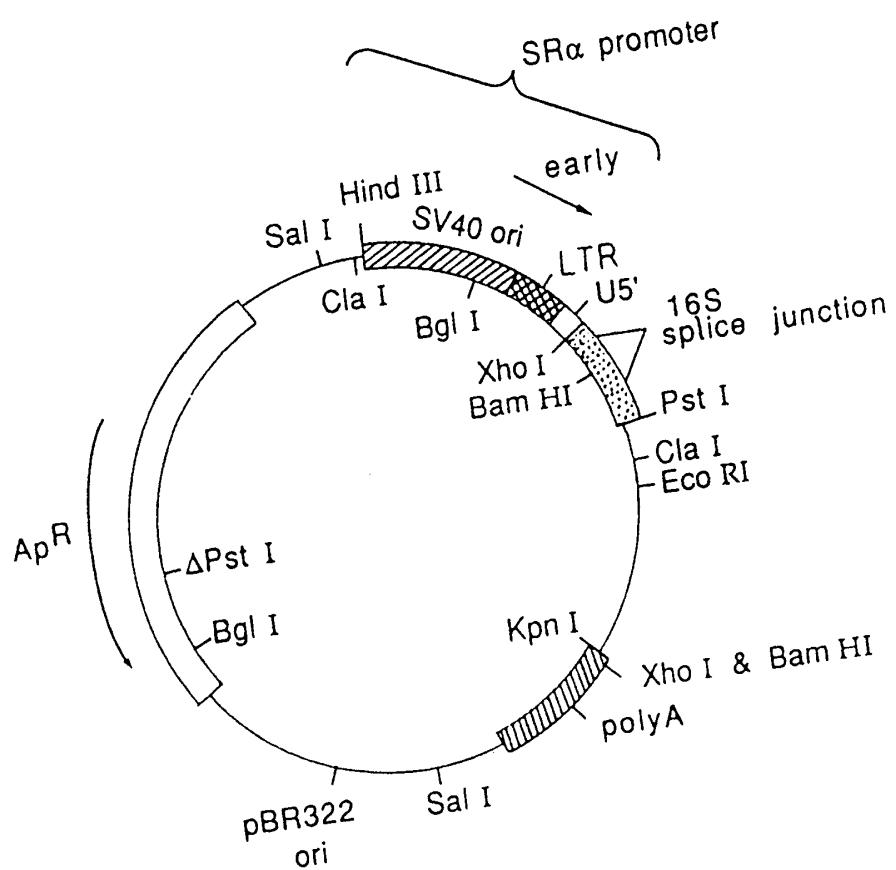


Figure 3

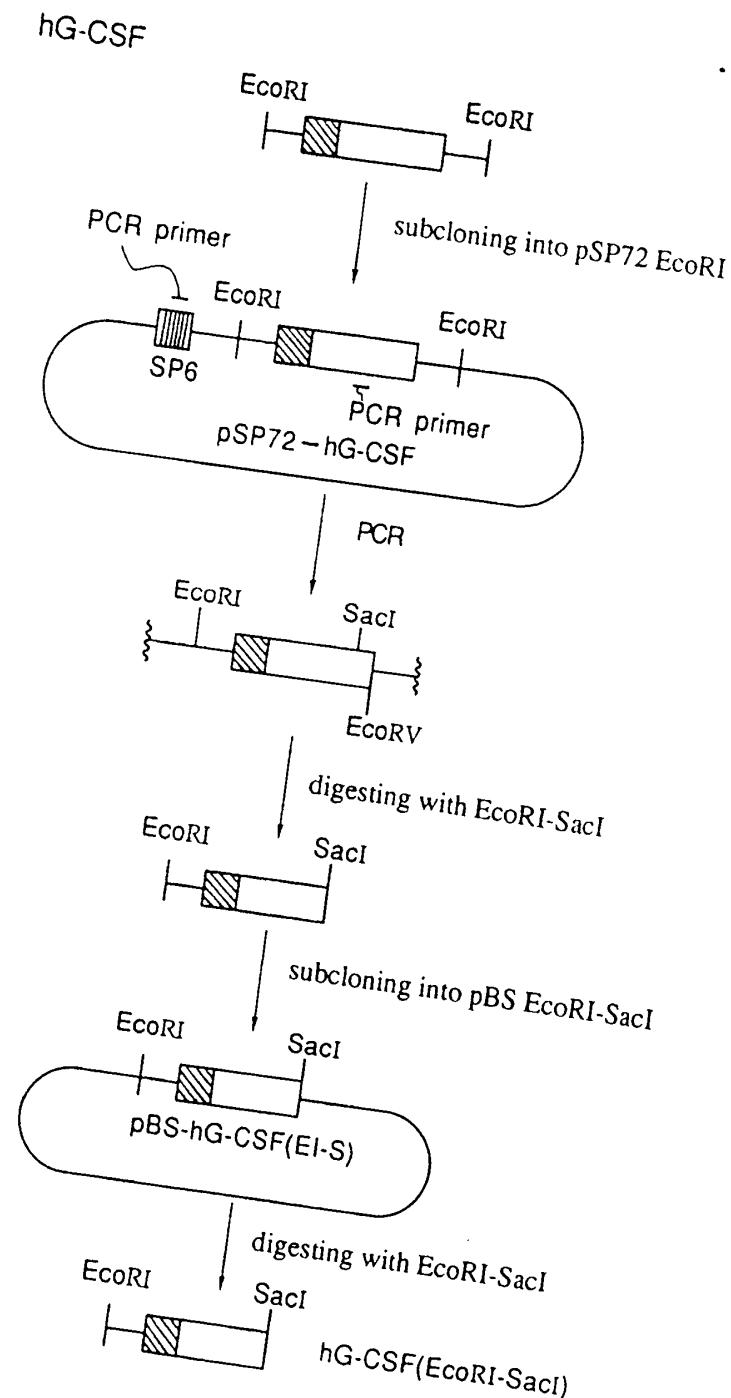


Figure 4

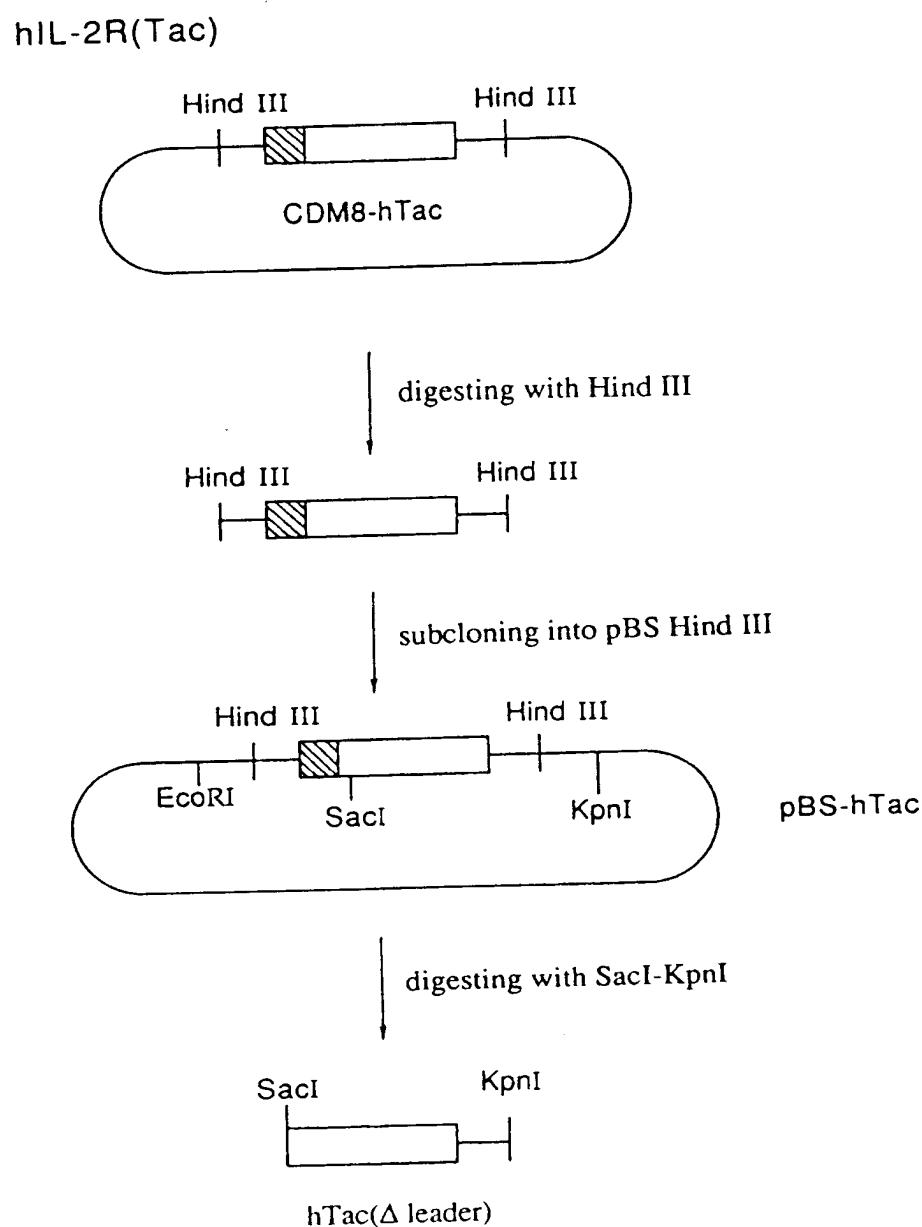
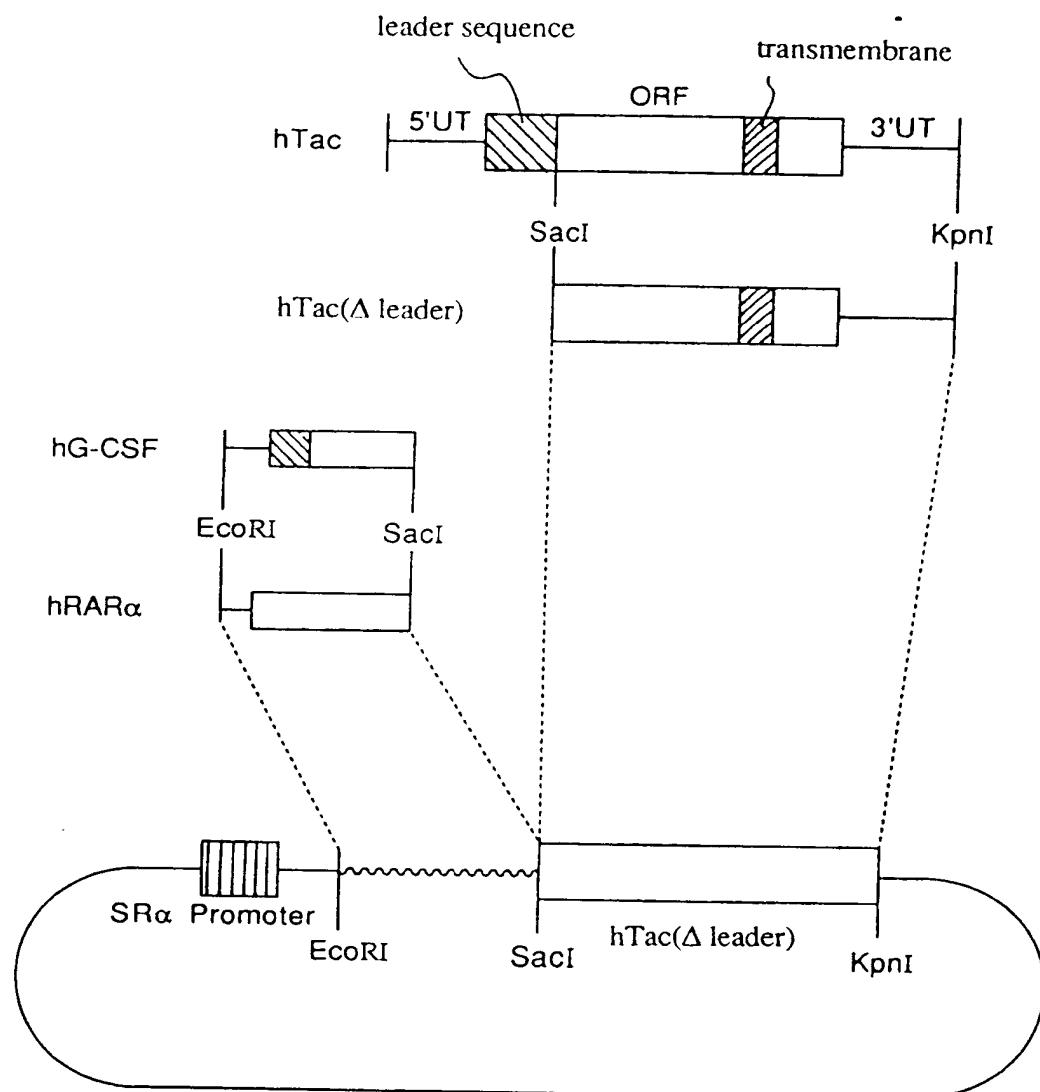


Figure 5



pcDL-SR α -hG-CSF-hTac(Δ leader) (pSGT)

pcDL-SR α -hRAR α -hTac(Δ leader) (pSRT)

Figure 6

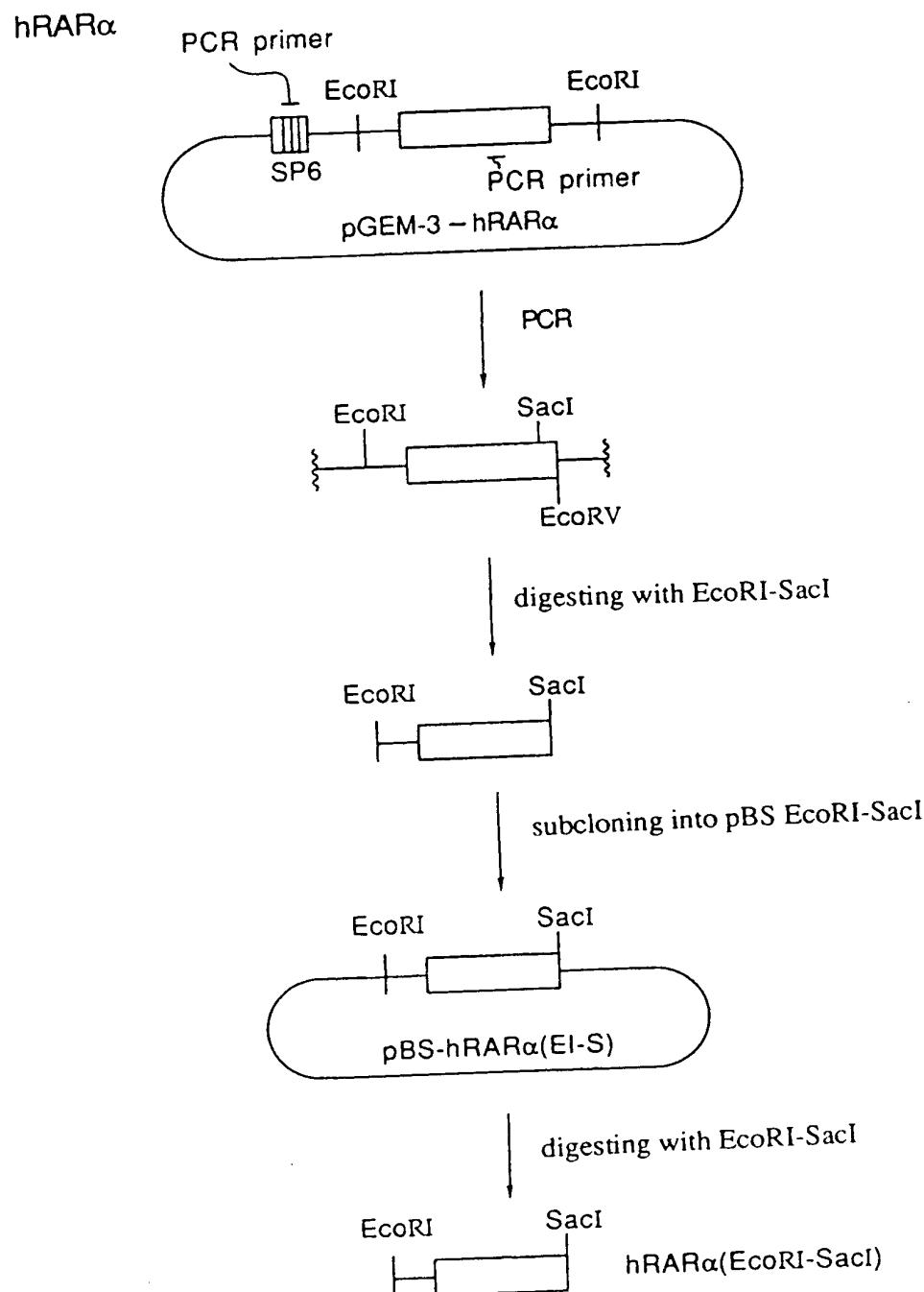


Figure 7

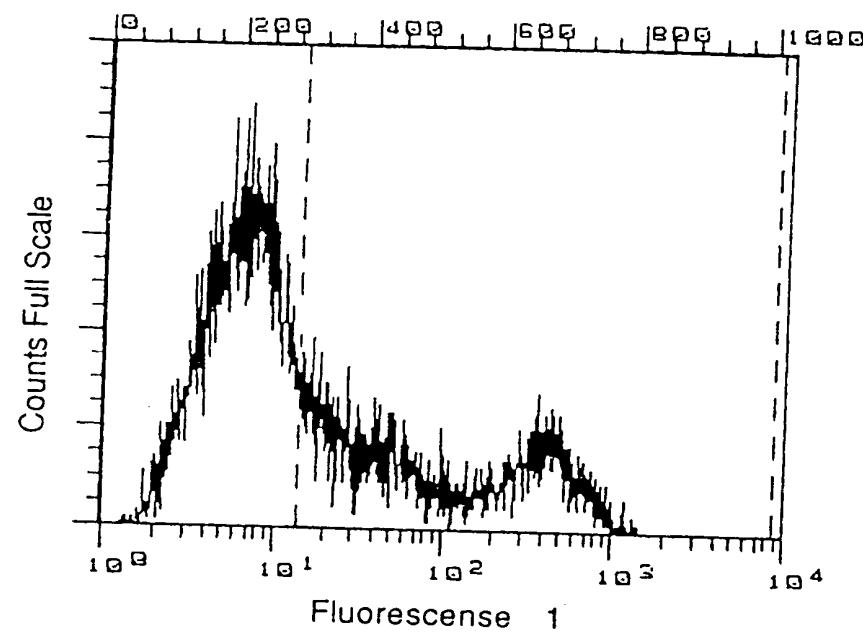


Figure 8

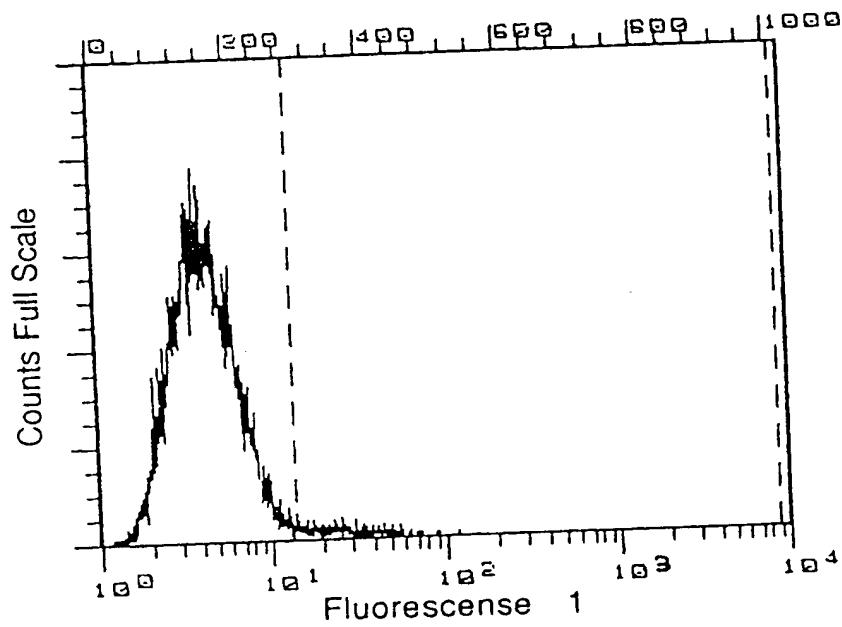


Figure 9

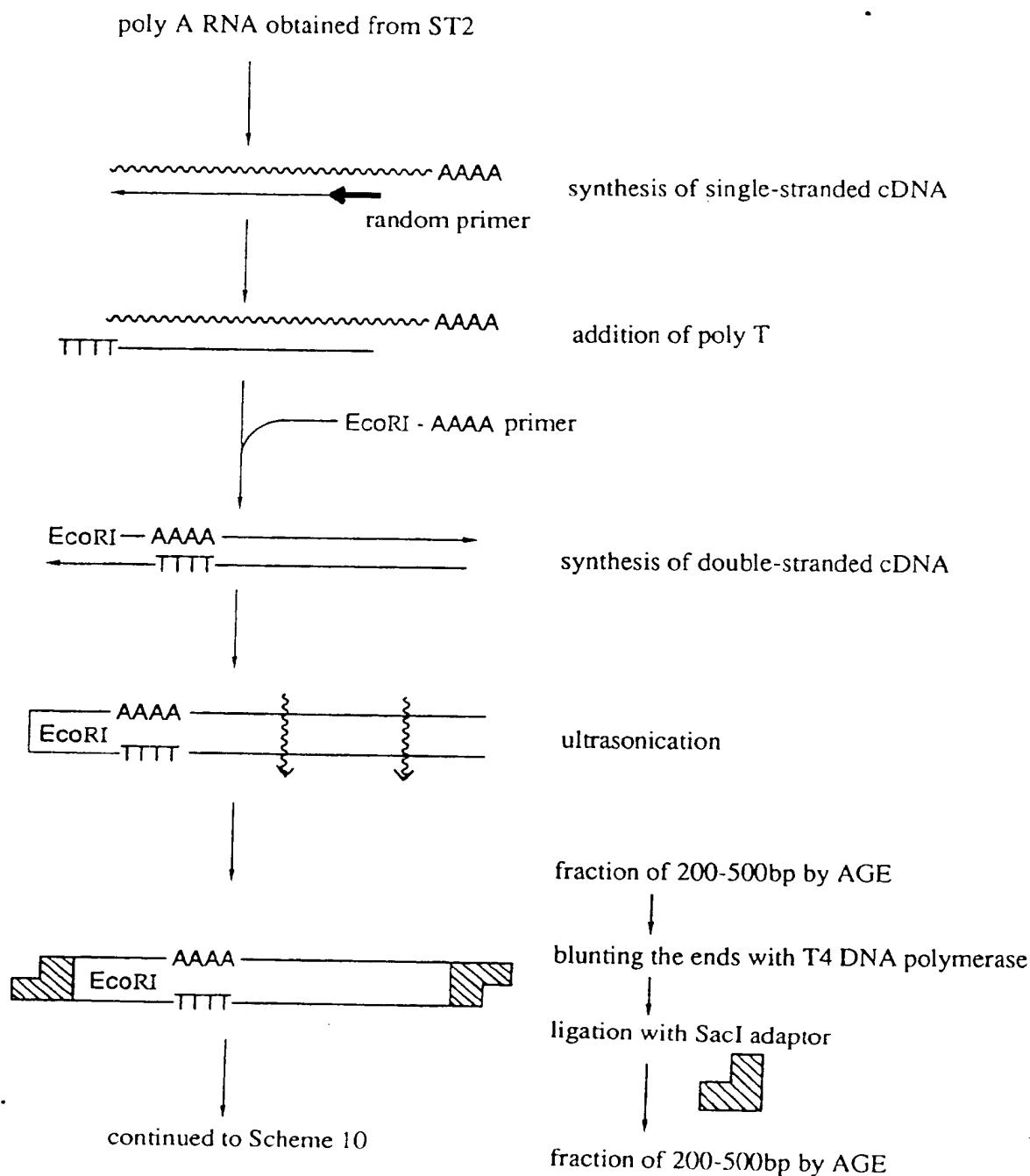


Figure 10

continued from Scheme 9

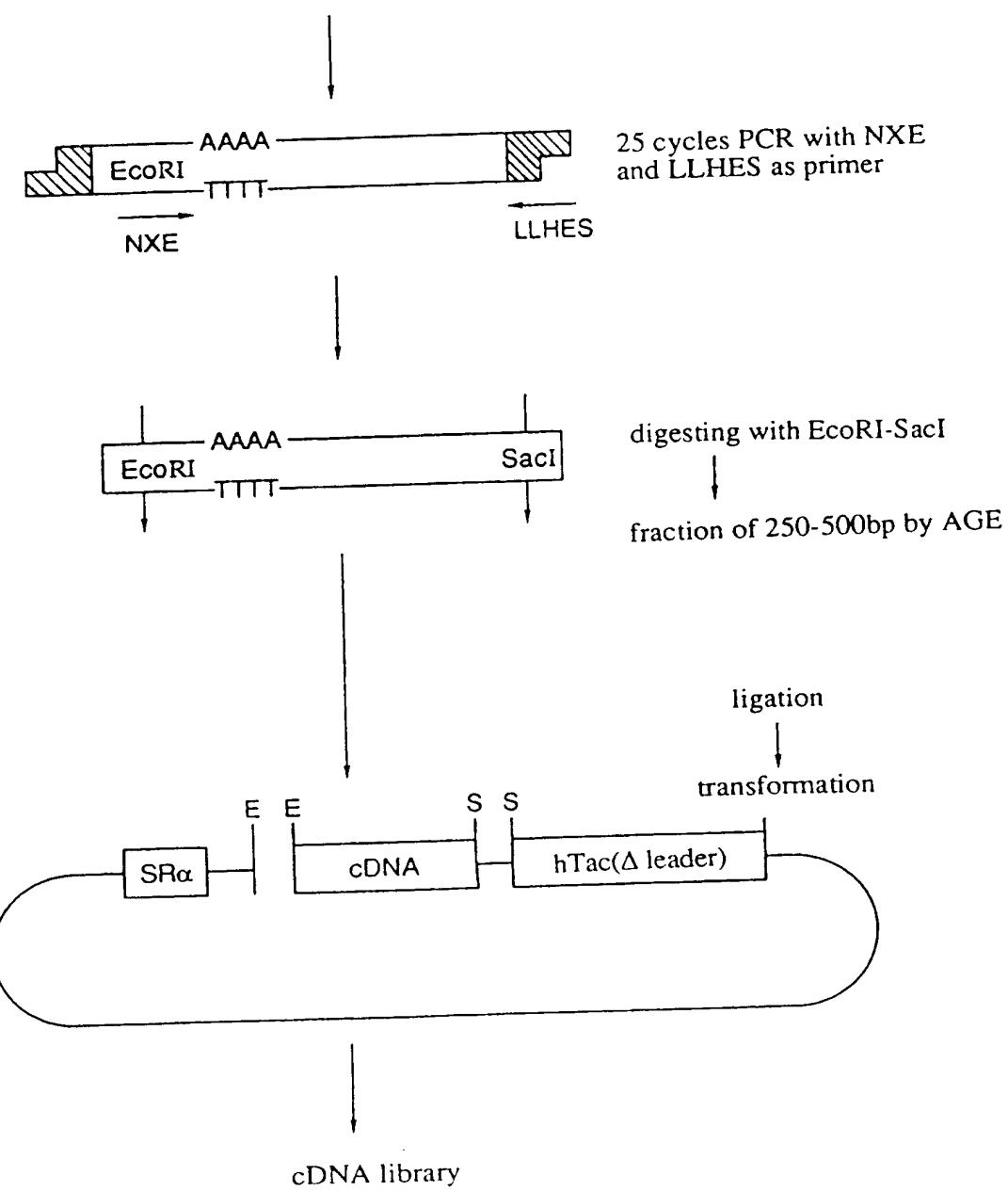
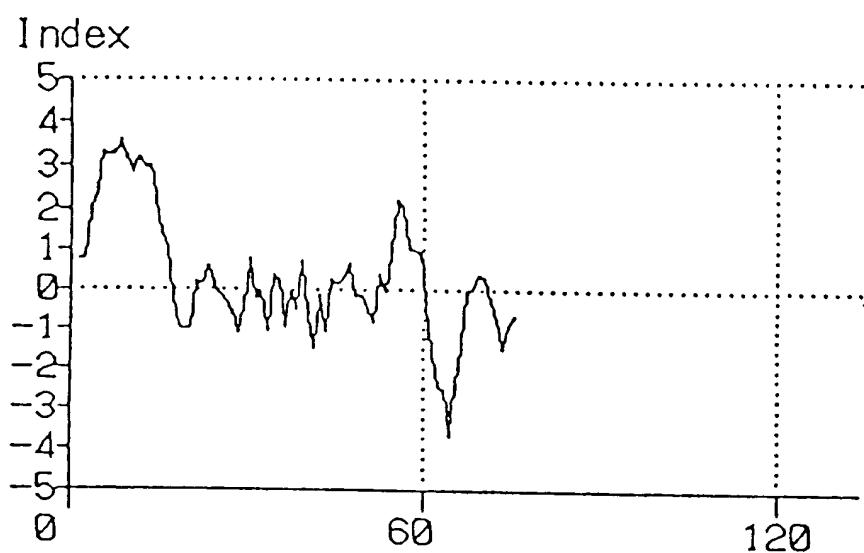


Figure 11





(12)

EUROPEAN PATENT APPLICATION

(21) Application number : 94300267.5

(22) Date of filing : 14.01.94

(51) Int. Cl.⁵ : C12N 15/10, C12N 15/12,
C12N 15/62, C12N 15/85,
C07K 13/00, C07K 15/28,
C12Q 1/68, A61K 37/02,
A61K 39/395

(30) Priority : 14.01.93 JP 22098/93

(43) Date of publication of application :
20.07.94 Bulletin 94/29

(84) Designated Contracting States :
AT BE CH DE DK ES FR GB GR IE IT LI LU MC
NL PT SE

(88) Date of deferred publication of search report :
08.02.95 Bulletin 95/06

(71) Applicant : Honjo, Tasuku
Kanyu-chi,
Oiwake-cho
Kita-Shirakawa, Sakyo-ku, Kyoto (JP)

(71) Applicant : ONO PHARMACEUTICAL CO., LTD.
1-5, Doshimachi 2-chome
Chuo-ku Osaka 541 (JP)

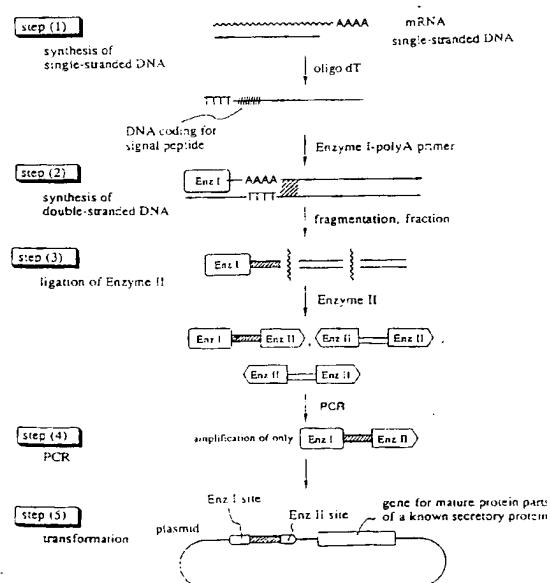
(72) Inventor : Honjo, Tasuku Kanyu-chi
Oiwake-cho
Kita-Shirakawa
Sakyo-ku Kyoto (JP)
Inventor : Tashiro, Kei
93 Koyama Higashi
Ohno-cho
Kita-ku
Kyoto-shi Kyoto (JP)
Inventor : Tade, Hideaki c/o Minase Research
Institute
Ono Pharmaceuticals Co.Ltd.
3-1-1, Sakurai
Shimamoto-cho Mishima-gun Osaka (JP)

(74) Representative : Bentham, Stephen et al
J.A. KEMP & CO.
14 South Square
Gray's Inn
London WC1R 5LX (GB)

(54) Novel process for constructing a cDNA library and a novel polypeptide and DNA coding the same.

(57) A process for constructing a cDNA library which has a selectivity for signal peptides, that makes it possible to efficiently identify unknown and useful polypeptide comprising a signal peptide. A novel polypeptide consisting of 89 amino acids (including a signal peptide) produced by a stroma cell line, which is useful as an agent for preventing or treating, for example, anemia, leukopenia or infections and the like, and DNAs coding for said polypeptide, have been identified using the process of the invention.

Figure 1





European Patent
Office

EUROPEAN SEARCH REPORT

Application Number

EP 94 30 0267

DOCUMENTS CONSIDERED TO BE RELEVANT					
Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	CLASSIFICATION OF THE APPLICATION (Int.Cl.5)		
Y	<p>NUCL. ACIDS RES., vol.19, no.25, December 1991, IRL PRESS, OXFORD, ENGLAND; pages 7097 - 7104</p> <p>E. HARA ET AL. 'Subtractive cDNA cloning using oligo(dT)30-latex and PCR: isolation of cDNA clones specific to undifferentiated human embryonal carcinoma cells'</p> <p>* page 7098, left column, line 12 - page 7099, left column, line 15; figure 2 *</p> <p>---</p>	1	C12N15/10 C12N15/12 C12N15/62 C12N15/85 C07K13/00 C07K15/28 C12Q1/68 A61K37/02 A61K39/395		
Y	<p>WO-A-92 18521 (LIFE TECHNOLOGIES, INC.) 29 October 1992</p> <p>* figures 4,5 *</p> <p>---</p>	1			
Y	<p>EP-A-0 244 042 (GIST-BROCADES N.V.) 4 November 1987</p> <p>* the whole document *</p> <p>---</p>	1			
Y	<p>DE-A-39 01 681 (BEHRINGERWERKE AG) 26 July 1990</p> <p>* claims 6-11 *</p> <p>---</p>	1			
P, X	<p>SCIENCE, vol.261, 30 July 1993, AAAS, WASHINGTON, DC, US; pages 600 - 603</p> <p>K. TASHIRO ET AL. 'Signal sequence trap: A cloning strategy for secreted proteins and type I membrane proteins'</p> <p>* page 600, middle column, line 1 - page 602, right column, line 3 *</p> <p>---</p> <p>-/-</p>	1-14	TECHNICAL FIELDS SEARCHED (Int.Cl.5) C12N C12Q C07K A61K		
The present search report has been drawn up for all claims					
Place of search	Date of completion of the search	Examiner			
THE HAGUE	15 December 1994	Hornig, H			
CATEGORY OF CITED DOCUMENTS					
X : particularly relevant if taken alone Y : particularly relevant if combined with another document of the same category A : technological background O : non-written disclosure P : intermediate document					
T : theory or principle underlying the invention E : earlier patent document, but published on, or after the filing date D : document cited in the application L : document cited for other reasons & : member of the same patent family, corresponding document					



European Patent
Office

EUROPEAN SEARCH REPORT

Application Number
EP 94 30 0267

DOCUMENTS CONSIDERED TO BE RELEVANT									
Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	CLASSIFICATION OF THE APPLICATION (Int.Cl.5)						
T	<p>PROC. NATL. ACAD SCI., vol.91, no.6, 15 April 1994, NATL. ACAD SCI., WASHINGTON, DC, US; pages 2305 - 2309</p> <p>T. NAGASAWA ET AL. 'Molecular cloning and structure of a pre-B-cell growth-stimulating factor' * page 2308, right column, line 17 - page 2308, left column, line 48; figure 3 *</p> <p>-----</p>	3-14							
TECHNICAL FIELDS SEARCHED (Int.Cl.5)									
<p>The present search report has been drawn up for all claims</p> <table border="1" style="width: 100%; border-collapse: collapse;"> <tr> <td style="width: 33%;">Place of search</td> <td style="width: 33%;">Date of completion of the search</td> <td style="width: 33%;">Examiner</td> </tr> <tr> <td>THE HAGUE</td> <td>15 December 1994</td> <td>Hornig, H</td> </tr> </table>				Place of search	Date of completion of the search	Examiner	THE HAGUE	15 December 1994	Hornig, H
Place of search	Date of completion of the search	Examiner							
THE HAGUE	15 December 1994	Hornig, H							
<p>CATEGORY OF CITED DOCUMENTS</p> <p>X : particularly relevant if taken alone Y : particularly relevant if combined with another document of the same category A : technological background O : non-written disclosure P : intermediate document</p> <p>T : theory or principle underlying the invention E : earlier patent document, but published on, or after the filing date D : document cited in the application L : document cited for other reasons & : member of the same patent family, corresponding document</p>									

